Relatório de CT-213: Deep Q-Learning

Henrique F. Feitosa

Instituto Tecnológico de Aeronáutica,
São José dos Campos, São Paulo, Brasil

1 Introdução

Nessa prática, buscou-se resolver o problema de *Mountain car* usando *Deep Q-Learning*. Inicialmente, implementou-se a arquitetura da rede que está descrita na figura 1. Após isso, implementaram-se uma escolha de ação usando uma política $\varepsilon - greedy$ e uma heurística chamada *reward engineering*, a qual pode ser descrita pelas equações 1 e 2.

Layer	Neurons	Activation Function
Dense	24	ReLU
Dense	24	ReLU
Dense	`action_size`	Linear

Figura 1. Mostra a arquitetura da rede neural a ser implementada.

$$r_{modified} = r_{original} + (position - start)^2 + (velocity)^2 \tag{1}$$

$$r'_{modified} = r_{modified} + 50 \cdot 1\{next - position \geq 0.5\} \tag{2}$$

Desssa forma, com a implementação feita, a rede foi treinada executando-se 300 episódios e , no final do treinamento, foi obtido um gráfico de retorno da tarefa ao longo dos episódios.

Finalmente, foi feita uma avaliação de política com 30 casos e anotou-se a taxa de sucesso no final da avaliação.

2 Resultados e Discussão

A arquitetura da rede neural implementada está na figura 2.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 24)                72
_________________________________________________________________
dense_2 (Dense)              (None, 24)                600
_________________________________________________________________
dense_3 (Dense)              (None, 3)                 75
=================================================================
Total params: 747
Trainable params: 747
Non-trainable params: 0
```

Figura 2. Mostra a arquitetura da rede neural implementada.

Comparando as figuras 1 e 2, pode-se perceber que as duas arquiteturas estão coerentes.

Após isso, o resultado do treinamento está representado pela figura 3, que explicita o retorno da tarefa ao longo dos episódios.

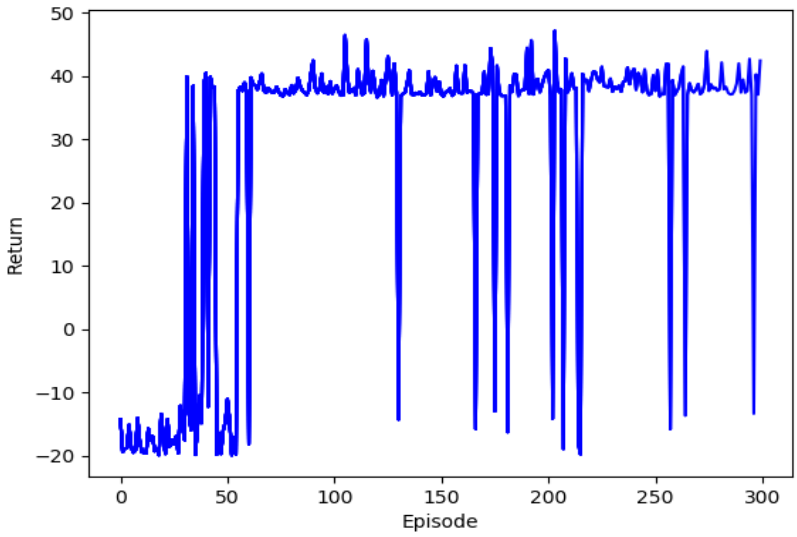

Figura 3. Mostra o retorno da tarefa ao longo dos episódios no treinamento.

Assim, pode-se perceber pela figura 3 que a partícula conseguiu atingir seu objetivo algumas vezes antes do episódio 100 e a maioria das vezes até o episódio 300, o que mostra que o treinamento foi eficiente e teve um bom retorno.

Finalmente, os resultados da avaliação de política estão representados pelas figuras 4 e 5.

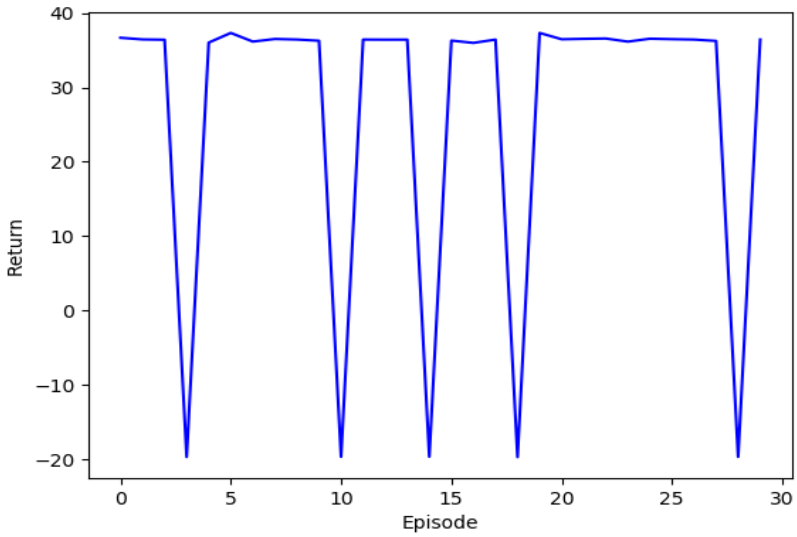

Figura 4. Mostra o retorno ao longo dos episódios

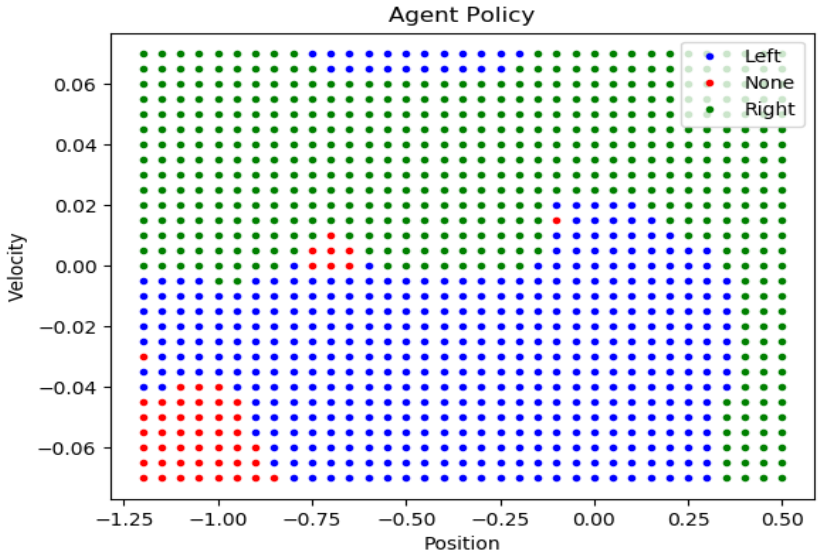

Figura 5. Mostra a política ótima encontrada.

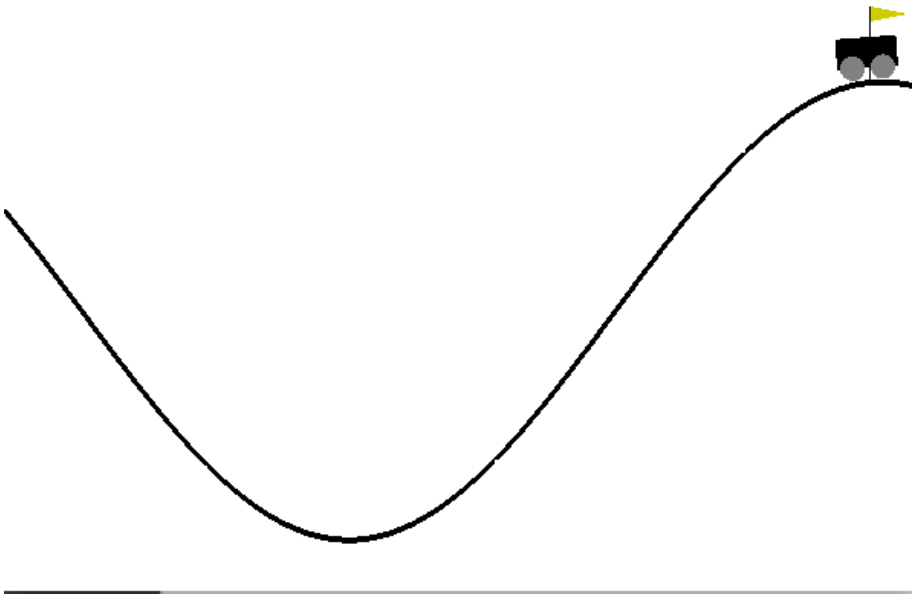

Figura 6. Mostra um episódio em que o desafio foi concluido.

Finalmente, obteve-se que a taxa de sucesso foi de 83.3%, o que demontra um desempenho satisfatório, visto que é maior que 70.0%.